Pathways of Maximum Likelihood for Rare Events in Nonequilibrium Systems: Application to Nucleation in the Presence of Shear

Matthias Heymann[1] and Eric Vanden-Eijnden[2]

[1]*Duke University Mathematics Department, Durham, North Carolina 27708, USA*

[2]*Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA*

(Received 31 August 2007; published 9 April 2008)

Even in nonequilibrium systems, the mechanism of rare reactive events caused by small random noise is predictable because they occur with high probability via their maximum likelihood path (MLP). Here a geometric characterization of the MLP is given as the curve minimizing a certain functional under suitable constraints. A general purpose algorithm is also proposed to compute the MLP. This algorithm is applied to predict the pathway of transition in a bistable stochastic reaction-diffusion equation in the presence of a shear flow, and to analyze how the shear intensity influences the mechanism and rate of the transition.

DOI: 10.1103/PhysRevLett.100.140601 PACS numbers: 05.10.−a, 05.40.−a, 05.70.Fh, 05.70.Ln

The description of rare reactive events in nonequilibrium systems which lack detailed balance represents a difficult theoretical and computational challenge. Examples of such events include phase transitions, chemical reactions, biochemical switches, or regime changes in climate. Brute force simulation of these events by Monte Carlo [1] or direct simulation of Langevin equations [2] is difficult because of the huge disparity between the time step which must be used to perform the simulations and the time scale on which the rare events occur. Familiar concepts such as the minimum energy path which are often used to explain rare events are inappropriate because there is no energy landscape over which the system navigates. Methods to accelerate the simulation of these events [3] also typically fail in nonequilibrium situations. Despite recent theoretical [4] and numerical [5] advances to bypass these difficulties, many issues remain open.

One aspect of rare events which can be exploited is that the pathway of these events is often predictable, even in nonequilibrium systems. This is the essence of large deviation theory [6]. When an improbable event occurs, the probability that it does so in any other way than the most likely one is very small because all these other ways are even much less probable. It then becomes important to identify the maximum likelihood path (MLP) of the rare event, i.e., the path which maximizes the event likelihood over all possible pathways and times this event can take to occur. The MLP explains the mechanism of the rare event, which is usually nontrivial and informative in complex systems. The MLP also permits to calculate other important quantities such as the rate of the event. In this Letter, we show that the MLP can be characterized geometrically as the directed curve minimizing a certain action functional, and we design an efficient general purpose algorithm to compute the MLP. We use this algorithm to study phase transitions in a Ginzburg-Landau-type model in the presence of shear [7]. In first approximation this model can be used as a crude approximation to describe, e.g., nucleation of protein crystals, a problem which has received lots of attention recently because of its importance for the pharmaceutical industry [8]. We find the MLP in this model and estimate the rate of the transition in function of the shear intensity. We also show how to automatically identify the shear intensity which maximizes this rate. Part of the results presented here are justified rigorously in Ref. [9].

The path of maximum likelihood.—Consider a dynamical system whose evolution is governed by the following stochastic differential equation (SDE):

$$x'(s) = b(x(s)) + \sqrt{\varepsilon}\,\eta(s). \tag{1}$$

Here $x(s) \in \mathbb{R}^n$ specifies the state of the system at time s, $x'(s) = dx/ds$ (the reason we use s rather t to denote time and prime rather than dot to denote the time derivative will become clear shortly), $b(x)$ is the deterministic drift, and $\sqrt{\varepsilon}\,\eta(s)$ is a noise term of square amplitude ε, $\eta(s)$ being a Gaussian white-noise with mean 0 and covariance $\langle \eta_i(s)\eta_j(s')\rangle = \delta_{ij}\delta(s - s')$. For simplicity we assumed that the noise is additive in (1) (i.e., the diffusion coefficient is constant), but this assumption is unessential and will be relaxed below.

Suppose that the deterministic system associated with (1) when $\varepsilon = 0$, i.e., $x'(s) = b(x(s))$, has two stable equilibrium points, x_1 and x_2. This means that, in the absence of noise, any trajectory starting close to x_1 or x_2 is attracted by this point and remains there forever once it has reached it. Any amount of noise $\varepsilon > 0$ changes this picture. With the noise, even if the trajectory starts at x_1, say, it will eventually make a transition to the vicinity of x_2. We are interested in understanding how these transitions occur when the noise amplitude is small, $0 < \varepsilon \ll 1$.

To answer this question, recall that the probability density of the solution $x(s)$ of (1) with $x(0) = x_1$ and $x(T) = x_2$ is formally proportional to $e^{-S_T(x)/\varepsilon}$, where

$$S_T(x) = \frac{1}{2}\int_0^T |x'(s) - b(x(s))|^2 ds. \quad (2)$$

When $\varepsilon \ll 1$, this probability density is very peaked around the path $x^\star(s)$ which maximizes $e^{-S_T(x)/\varepsilon}$, i.e., the path which minimizes the action $S_T(x)$ over all paths $x(s)$ subject to $x(0) = x_1$ and $x(T) = x_2$ *and* over all $T > 0$. We call this path the path of maximum likelihood for the transition. Note that the minimization over T is consistent with viewing $e^{-S_T(x)/\varepsilon}$ as a probability density for both x and T, and that this minimization is crucial to obtain the path whose likelihood is maximal over all paths going from x_1 to x_2 in any possible time. The double minimization over x and T turns out to be problematic, however, because the minimum is reached when $T \to \infty$ in general. Thus there is some intrinsic difficulty to define the MLP if we work with physical time. We can, however, rescale time locally to get a geometric characterization of the MLP in which this problem disappears.

To do this, first use $|x'|^2 + |b(x)|^2 \geq 2|x'||b(x)|$ to deduce that $\min_{x,T} S_T(x) \geq \min_{x,T} \hat{S}_T(x)$, with

$$\hat{S}_T(x) = \int_0^T \{|x'||b(x)| - (x', b(x))\} ds, \quad (3)$$

where $(x', b(x))$ denotes the dot product of x' and $b(x)$. Now for two key observations. First, the inequality $\min_{T,x} S_T(x) \geq \min_{T,x} \hat{S}_T(x)$ is actually an equality, $\min_{T,x} S_T(x) = \min_{T,x} \hat{S}_T(x)$, because time can be rescaled locally along $x(s)$ without changing the geometric location of $x(s)$ but in such a way that $|x'(s)| = |b(x(s))|$: in this case, $|x'|^2 + |b(x)|^2 = 2|x'||b(x)|$. Second, $\min_{x,T} \hat{S}_T(x) \equiv \min_x \hat{S}_T(x)$ because $\min_x \hat{S}_T(x)$ is independent of T. Indeed, $\hat{S}_T(x)$ is a line integral along the directed curve $\gamma = \{x(s): s \in [0, T]\}$, which only depends on the curve γ but not on the way γ is parametrized. This means that we can just as well set $T = 1$ in (3) and concentrate on minimizing this action over all $x(s)$ such that $x(0) = x_1$ and $x(1) = x_2$, remembering that the "time" s in these $x(s)$ is not the physical time anymore but simply a parameter to localize points along these paths (which finally justifies our unusual choice of s to denote time). Since only the geometric location of $\{x(s): s \in [0, 1]\}$ matters, we can use this flexibility to fix the rescaling of time along $x(s)$, e.g., by imposing that $|x'(s)| = \text{cst}$, a choice which corresponds to interpreting s as normalized arclength along the path and will be convenient in the numerics. With this choice, we are then left with the following minimization problem

$$\min_{\substack{x(s) \\ x(0)=x_1, x(1)=x_2}} \hat{S}_1(x) \quad \text{subject to } |x'(s)| = \text{cst}. \quad (4)$$

The minimizer $x^\star(s)$ of (4) is the MLP, and $e^{-\hat{S}_1(x^\star)/\varepsilon}$ gives, up to a prefactor, the transition rate [9].

Numerical aspects.—The geometric reformulation leading to the minimization problem in (4) is particularly convenient for numerical purposes. Recall that the minimizer of (4) must satisfy the Euler-Lagrange equation, $\delta\hat{S}_1/\delta x = 0$. A straightforward calculation from (3) shows that $\delta\hat{S}_1/\delta x$ satisfies

$$\lambda \delta\hat{S}_1/\delta x = -\lambda^2 x'' + \lambda(\nabla b - \nabla b^T)x' + (\nabla b)^T b - \lambda\lambda' x', \quad (5)$$

where $\lambda = |b|/|x'|$, ∇b is the tensor with entries $\partial b_i/\partial x_j$ and ∇b^T denotes its transpose. To solve $\delta\hat{S}_1/\delta x = 0$ we use a relaxation method which is equivalent to minimizing (4) by a preconditioned steepest-descent algorithm in the direction of $-\lambda\delta\hat{S}/\delta x$ [10]. We start from a path $x^0(s)$ with $x^0(0) = x_1$ and $x^0(1) = x_2$ discretized into $N + 1$ points, $x_i^0 = x^0(i/N)$, $i = 0, \ldots, N$, and for $k \geq 0$ update these points till convergence by iterating on: (i) Given $\{x_i^k\}_{i=0,\ldots,N}$, set $\lambda_0^k = \lambda_N^k = 0$ and compute $x_i'^k = (x_{i+1}^k - x_{i-1}^k)/(2/N)$, $\lambda_i^k = |b(x_i^k)|/|x_i'^k|$, and $\lambda_i'^k = (\lambda_{i+1}^k - \lambda_{i-1}^k)/(2/N)$ for $i = 1, \ldots, N-1$. (ii) Find $\tilde{x}_i$ for $i = 1, \ldots, N-1$ by solving

$$\frac{\tilde{x}_i - x_i^k}{\tau} = (\lambda_i^k)^2 \frac{\tilde{x}_{i+1} - 2\tilde{x}_i + \tilde{x}_{i-1}}{1/N^2} - \lambda_i^k[\nabla b(x_i^k) - \nabla b^T(x_i^k)]x_i'^k - \nabla b^T(x_i^k)b(x_i^k) + \lambda_i^k\lambda_i'^k x_i'^k \quad (6)$$

with boundary conditions $\tilde{x}_0 = x_1$ and $\tilde{x}_N = x_2$; $\tau > 0$ plays the role of the updating time-step. (iii) Interpolate a curve across $\{\tilde{x}_i\}_{i=0,\ldots,N}$ and discretize this curve to find $\{x_i^{k+1}\}_{i=0,\ldots,N}$ so that $|x_{i+1}^{k+1} - x_i^{k+1}| = |x_i^{k+1} - x_{i-1}^{k+1}|$ for $i = 1, \ldots, N-1$.

This algorithm has several pleasant features [9]. The semi-implicit updating in (6) guarantees that the code is stable with a time-step τ independent of N [11]; the tridiagonal system in (6) can be solved componentwise using the Thomas algorithm (see, e.g., Sec. 2.9 in [11]), and the reparametrization and interpolation in step (iii) used to enforce the discrete equivalent of the constraint $|x'| \equiv \text{cst}$ can be done in $O(N)$ operations as in the string method [12]. More sophisticated minimization procedures can be used, like, e.g., preconditioned conjugate gradient or multigrid methods, but the algorithm above proved sufficient for our needs. We called this algorithm the geometric minimum action method (gMAM) since it is a generalization of the minimum action method introduced in [13], adapted to the geometric action $\hat{S}_1(x)$.

Application to nucleation in the presence of shear.—As an illustration of our procedure, we compute the MLP in a stochastic reaction-diffusion equation in the presence of shear. Denoting by $\phi(x, y, t)$ the order parameter in the system, and going back using t to denote time and x and y for space to avoid confusions, this equation is

$$\dot{\phi} = \kappa\Delta\phi + \phi - \phi^3 + c\sin(2\pi y)\partial_x\phi + \sqrt{\varepsilon}\eta. \quad (7)$$

This equation is a perturbed version of the Ginzburg-Landau equation studied in [7] (see also [14]); we assume

periodic boundary conditions on $(x, y) \in \Omega \equiv [0, 1]^2$, the dot denotes the time derivative, $\Delta = \partial_x^2 + \partial_y^2$ is the Laplacian, $\kappa > 0$ is the diffusion coefficient, and $\eta(x, y, t)$ is a spatiotemporal white-noise with covariance $\langle \eta(x, y, t)\eta(x', y', t')\rangle = \delta(x - x')\delta(y - y')\delta(t - t')$ [15]. The term $c \sin(2\pi y)\partial_x \phi$ mimics the presence of a nonlinear shear flow with shear rate c in the x direction. Without this term (i.e., for $c = 0$), (7) is the perturbed steepest-descent dynamics on the Ginzburg-Landau energy functional $E(\phi) = \int_\Omega [\frac{1}{2}|\nabla\phi|^2 + V(\phi)]dxdy$ with potential $V(\phi) = \frac{1}{4}(1 - \phi^2)^2$. This energy is minimized by the two homogeneous states $\phi = \pm 1$ which are metastable states. The shear term $c \sin(2\pi y)\partial_x \phi$ does not destroy metastability of these two states, but it destroys the detailed balance of the dynamics. In other words, noise-induced transitions between $\phi = -1$ and $\phi = +1$ still occur, but they are nonequilibrium transitions. We want to compute the MLP of these transitions and estimate their rate as a function of the shear intensity c. As mentioned before, questions of this type are of importance, e.g., in the context of protein crystallization [8].

Our procedure and the gMAM algorithm can be straightforwardly applied to (7). First, we Fourier transform Eq. (7) using a pseudospectral representation for the terms $-\phi^3 + c \sin(2\pi y)\partial_x \phi$. Then we discretize this equation on a regular grid with 64^2 points on $[0, 1]^2$. The resulting system of equations is a large set of SDEs similar to (1). We use a MatLab implementation of gMAM with $N = 100$ and τ between 0.1 (for $c = 0$) and 0.002 (for $c = 1$) to compute the MLP at various c, see Fig. 1. We also compute the value of the action on the minimizer, $\hat{S}_1(\phi^\star)$, which is now a function of the shear intensity c and gives the estimate $e^{-\hat{S}_1(\phi^\star)/\varepsilon}$ for the rate of the transition at this shear intensity. This result is shown in Fig. 2, and together with the ones shown in Fig. 1 they indicate the following. A small amount of shear helps the nucleation. Instead of having to create a long horizontal droplet which reconnects by the sides, which is the MLP when $c = 0$ (top panel in Fig. 1), in the presence of shear ($c > 0$ but not too large), the system is able to nucleate a smaller and slightly rotated ellipsoidal droplet which is then expanded horizontally both by the noise and the shear flow itself (middle panel in Fig. 1). The fact that this mechanism facilitates the transitions is confirmed by the result in Fig. 2 which shows that $\hat{S}_1(\phi^\star)$ first decreases as c increases (and hence the rate $e^{-\hat{S}_1(\phi^\star)/\varepsilon}$ of the transition increases). Too much shear, however, stops being a benefit. Indeed, when c becomes too large, the nucleating droplet is quickly stretched horizontally before being able to reach a size sufficient to increase the likelihood of the transition (lower panel in Fig. 1). In this case, we essentially go back to the axisymmetric situation at $c = 0$, which is also confirmed by the asymptotic value reached by $\hat{S}_1(\phi^\star)$ when c is large. Note that these results are consistent with those obtained in Ref. [16] using Monte Carlo simulations in an Ising model similar to (7) in spirit. Quite interestingly, our method can be generalized to determine automatically the value of c at which the rate of the transition is maximized, $c^\star \approx 0.1183$. It suffices to compute the derivative of $S_1(\phi^\star)$ with respect to c, which can be done explicitly [17], and then use this derivative in a standard minimization algorithm in which gMAM enters as a subroutine. This is the procedure we used to determine $c^\star$, and it can also be used to maximize the rate of an event over more complicated control parameters than the scalar c in the present example [17].

Generalizations.—Our procedure and the gMAM algorithm can be easily generalized to situations where the noise is nonadditive, i.e., when (1) is replaced by

$$x'(s) = b(x(s)) + \sqrt{\varepsilon}\sigma(x(s))\eta(s). \tag{9}$$

The only thing that needs to be changed is the inner product one works with: in the action (3), instead of using the Euclidean inner product, (x', b), we now need $(x', b)_a \equiv \sum_{i,j=1}^n a_{ij}^{-1}(x(s))x_i'(s)b_j(x(s))$, where $a_{ij}^{-1}(x)$ denotes the

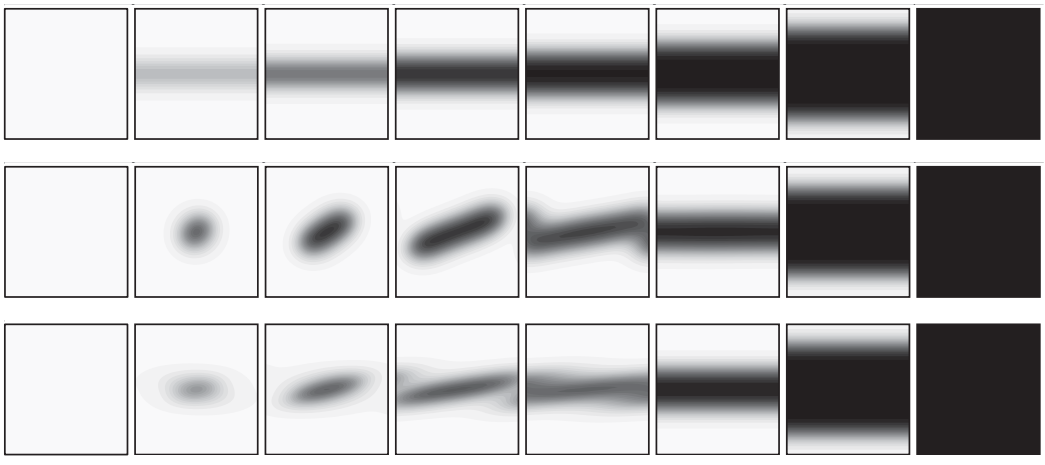

FIG. 1. Snapshots along three MLPs obtained at different values of the shear intensity: top panel, $c = 0$; middle panel, $c = c^\star \approx 0.1183$; bottom panel, $c = 0.6$. The MLP at $c = c^\star \approx 0.1183$ is the one which maximizes the rate of the transition over all shear intensities c. The color code goes from white $= -1$ to black $= 1$ and the transition from left to right.

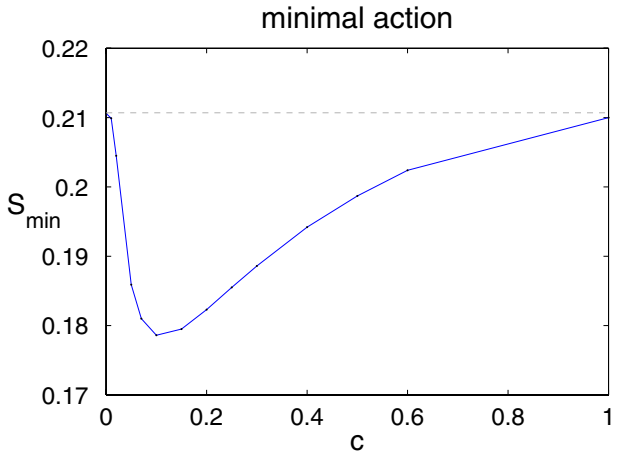

FIG. 2 (color online). The minimum along the MLP, $S_{\min} = \hat{S}_1(\phi^\star)$, as a function of the shear intensity c. A smaller $S_{\min}$ means a larger transition rate. $S_{\min}$ is minimum at $c^\star \approx 0.1183$, corresponding to the MLP shown in the middle panel of Fig. 1.

entries of the inverse of the matrix $a(x) = \sigma\sigma^T(x)$. Similarly, the norms $|x'|$ and $|b|$ are replaced by $(x', x')_a^{1/2}$ and $(b, b)_a^{1/2}$. The minimization of this action is performed by a generalization of gMAM [9].

Another interesting generalization is to continuous-time Markov jump processes in which there are J independent reaction channels with rates $\varepsilon^{-1}\nu_j(x)$ and which are such that the state of the system changes from x to $x + \varepsilon e_j$ when the jth reaction occurs. Our procedure and the gMAM algorithm can be generalized to handle these systems when ε is small; i.e., individual reactions are fast but the effect of each of them is small [9].

Concluding remarks.—The MLP of a rare event can be characterized geometrically as the curve minimizing the functional (3), and this leads to an efficient algorithm (gMAM) to compute it. These results apply to nonequilibrium transitions in systems which lack detailed balance in which the MLP results from a nontrivial interplay between the deterministic drift and the noise. In particular, unlike what happens with equilibrium transitions, in general the MLP of a nonequilibrium transition is not everywhere parallel or antiparallel to the drift b, and it is not reversible (i.e., the MLP from x_1 to x_2 is different from the one from x_2 to x_1). These nontrivial features were illustrated here by studying nucleation in the presence of shear, an example which made apparent the appealing features of our approach. gMAM permits to estimate efficiently and accurately the MLP of the rare event and even to optimize its rate over control parameters (e.g., the shear intensity in the example). Such calculations would be much more difficult to perform using direct numerical simulation of the stochastic systems.

We thank Giovanni Ciccotti and Jonathan Mattingly for useful comments. We also benefited from useful remarks from the referees. This work was partially supported by NSF grants Nos. DMS02-09959 and DMS02-39625, and by ONR grant No. N00014-04-1-0565.

[1] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975); D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[2] P. E. Kloeden and E. Platen, *The Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, 1992).
[3] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977); A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997); F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); P. G. Bolhuis, D. Chandler, C. Dellago, and P. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002); G. Bussi, A. Laio, and M. Parrinello, Phys. Rev. Lett. **96**, 090601 (2006).
[4] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997); J. Kurchan, J. Phys. A **31**, 3719 (1998); G. E. Crooks, Phys. Rev. E **60**, 2721 (1999); D. Evans and D. Searles, Adv. Phys. **51**, 1529 (2002).
[5] A. K. Faradjian and R. Elber, J. Chem. Phys. **120**, 10 880 (2004); R. J. Allen, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **94**, 018104 (2005); C. Giardina, J. Kurchan, and L. Peliti, Phys. Rev. Lett. **96**, 120603 (2006).
[6] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, New York, 1998), 2nd ed.; A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis - Queues, Communication, and Computing* (Chapman & Hall, London, 1995).
[7] A. Cavagna, A. J. Bray, and R. D. M. Travasso, Phys. Rev. E **62**, 4702 (2000).
[8] A. Chernov, J. Struct. Biol. **142**, 3 (2003); A. Penkova, O. Gliko, I. L. Dimitrov, F. V. Hodjaoglu, C. Nanev, and P. G. Vekilov, J. Cryst. Growth **275**, e1527 (2005); D. L. Chen, C. J. Gerdts, and R. F. Ismagilov, J. Am. Chem. Soc. **127**, 9672 (2005); A. Penkova, W. C. Pan, F. Hodjaoglu, and P. G. Vekilov, Ann. N.Y. Acad. Sci. **1077**, 214 (2006).
[9] M. Heymann, Ph.D. thesis, New York University, 2007; M. Heymann and E. Vanden-Eijnden, Commun. Pure Appl. Math. (to be published).
[10] We use $\lambda\delta\hat{S}_1/\delta x$ because, unlike $\delta\hat{S}_1/\delta x$, it remains finite at the critical points (i.e., at points x where $b(x) = 0$).
[11] K. W. Morton and D. F. Mayers, *Numerical Solution of Partial Differential Equations: An Introduction* (Cambridge University Press, Cambridge, England, 2005), 2nd ed.
[12] W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002); J. Chem. Phys. **126**, 164103 (2007).
[13] W. E, W. Ren, and E. Vanden-Eijnden, Commun. Pure Appl. Math. **57**, 637 (2004).
[14] A. Onuki, J. Phys. Condens. Matter **9**, 6119 (1997).
[15] This noise should be regularized in space in order that (7) be well posed, but we ignore this aspect since it only leads to higher-order corrections.
[16] C. Valeriani, Ph.D. thesis, AMOLF, 2007.
[17] M. Heymann and E. Vanden-Eijnden (to be published).